Email ☹

What can you learn (and not learn) by looking at it?

Where is it coming from?

How has it gotten so strange lately?

What can you do about it?

This talk was prompted by the recent surge in the total volume of SPAM email, largely due to the Klez worm and all its variants.

SPAM continues to *not* be a computer security issue in itself. If you get SPAM with criminal aspects you may, of course, report it to your supervisor, to computer_security, or directly to an appropriate law enforcement agency.

The Secret Life of Email

- The parts of an email message
  - The invisible: Envelope
  - The hidden: Headers
  - The inscrutable: (MIME) Body

- Transfer of an email message
  - User Agent to Transfer Agent
  - Transfer Agent to Transfer Agent
  - Transfer Agent to Delivery

Envelope

- This is how one Agent tells another about the sender and recipients.

```
220 receiving.agent.net
HELO sending.mailer.org
250 receiving.agent.net
MAIL From:<supposed.sender@some.host.org>    ⇦ Alleged
250 Sender OK
RCPT To:<actual.recipient@other.site.com>    ⇦ Actual
250 Recipient OK
DATA
354 Enter message.  End with "."
{ Headers and body sent here. ... No         ⇦ Arbitrary
 necessary correlation with the envelope. }
.
250 Message accepted
QUIT
221 Bye
```

For the most part, the sending process only needs to look at the three−digit codes which begin each reply from the receiving SMTP process.

The important point to notice here is that the entire message, including headers and body, is transferred after the DATA command and need not have any specific relationship to the sender and recipient information which have already been sent to the receiving process.

Envelope to Headers?

- Is envelope information reflected in the headers?
  - Nope. Sorry.
- Well, sometimes ... some of it ... by the deliberate action of Transfer Agents or Delivery Agents.
  - The envelope recipient may be shown in "Received:" headers.
  - The Delivery Agent often places the envelope sender address in a "Return−Path:" header.
- Of course, anything placed in the headers before leaving the last mailer you can't trust could be a complete fiction.
  - Including From:, To:, Cc:, Date:, Sender:, Message−id:, X−Authenticated−Sender: ... and previous Received: lines.

The Fermilab mail gateways do reflect the envelope recipient information in the Received: headers they add.

Acceptable Fictions

- Should the envelope sender be forced to be the same as the From: address in the header?
  - No. The envelope sender receives reports of delivery errors. A mailing list server, for example, might set the outgoing envelope sender to the list owner or an automated error−handling mailbox.
- Should the header recipient(s) be forced to include the envelope recipient(s)?
  - No. The current envelope list may not be complete. The recipient may want a clear distinction between mail from a list and individually−addressed mail. The sender may not want all recipients to see each others' addresses.

In other words, the logical independence of envelope and header information is not a bug, it's a feature. Don't look for it to be changed.

Sample 1 (part 1)

```
Return-Path: gjackson@uchicago.edu
Delivery-Date: Tue, 21 Aug 2001 13:16:51 -0500
Received: from heffalump.fnal.gov (heffalump.fnal.gov [131.225.9.20])
 by gungnir.fnal.gov (8.10.2+Sun/8.10.2) with ESMTP id f7LIGom19473
 for <crawdad@gungnir.fnal.gov>; Tue, 21 Aug 2001 13:16:50 -0500 (CDT)
Received: from CONVERSION-DAEMON.smtp.fnal.gov by smtp.fnal.gov
 (PMDF V6.0-24 #37519) id <0GIF00201K41MI@smtp.fnal.gov> for
 crawdad@gungnir.fnal.gov (ORCPT crawdad@fnal.gov); Tue,
 21 Aug 2001 13:16:51 -0500 (CDT)
```

- Taking the header lines from the top downward ...

- This message was delivered to a Unix mailbox, so the final mailer put the envelope source address into what's called a "Unix From line" or a "From-space" line. The message delimiter in such a file is "\nFrom ". Exmh turned that line into a Return-Path line and added the Delivery-Date.
- I trust my system, gungnir, to put its own Received line first, so I know this came from heffalump, and the envelope recipient on that transfer was crawdad@gungnir.fnal.gov.
- I trust smtp a/k/a heffalump, so I believe that it turned the original recipient ("ORCPT") crawdad@fnal.gov to crawdad@gungnir.fnal.gov. (See also next line ...)

This is the first of three parts of a legitimate email message. The dissection of it and the Klez worm message that follows shows what little information is available to distinguish a forgery.

Sample 1 (part 2)

```
Received: from suspect.uchicago.edu ([128.135.248.223])
 by smtp.fnal.gov (PMDF V6.0-24 #37519)
 with ESMTP id <0GIF00IHBK41OE@smtp.fnal.gov> for crawdad@gungnir.fnal.gov
 (ORCPT crawdad@fnal.gov); Tue, 21 Aug 2001 13:16:49 -0500 (CDT)
Received: from agh195.aps.anl.gov [164.54.89.195] by suspect.uchicago.edu
 with SMTPBeamer v3.25 ; Tue, 21 Aug 2001 13:16:45 -0500
Content-return: prohibited
Date: Tue, 21 Aug 2001 13:16:35 -0500
From: John Q Public <jqpublic@uchicago.edu>
```

- smtp.fnal.gov tells me that it received this message from a system which claimed (in the HELO command) to be called suspect.uchicago.edu and which had the IP address 128.135.248.223.
- suspect.uchicago.edu is outside my sphere of trust, but if the next line isn't faked, the message got there from a host claiming to be in the Argonne Guest House (agh195...) and with an IP address of 164.54.89.195. According to whois.arin.net his netblock really is associated with the Advanced Photon Source (aps) and Argonne, so all is well.
- The sending mailer has declared that if this message bounces, the body of the message should not be included in the returned error.
- The Date header can be supplied by the originator or will often be filled in by any intermediate host that finds it absent.
- Notice that the From address is not in the domain where the mail really originated. You can see that this is not an abnormal condition.

The Content-Return: header instructs downstream mailers that, in the event of a delivery failure, the content (body) is not to be returned to the sender.

Sample 1 (part 3)

```
Subject: lunch outline
To: jqpublic@uchicago.edu, crawdad@fnal.gov, j-doe@uchicago.edu,
   stan.ford@stanford.edu, mgstanley@lbl.gov, baggins1@llnl.gov,
   ed.macmahon@pnl.gov, begleyeljr@ornl.gov, rick.danko@jlab.org
Message-id: <DKEKJMGKGHHLNBKBADMHOEBJCAAA.jqpublic@uchicago.edu>
MIME-version: 1.0
X-MIMEOLE: Produced By Microsoft MimeOLE V5.50.4807.1700
X-Mailer: Microsoft Outlook IMO, Build 9.0.2416 (9.0.2911.0)
Content-type: multipart/mixed;
   boundary="Boundary_(ID_2rL0+Wg0ZG0rOtgVqB+e9g)"
Importance: Normal
X-Priority: 3 (Normal)
X-MSMail-priority: Normal
```

- I have no proof that the message actually went to any of the other addresses.
- The Message-id should end with a full domain name and be unique.
- Headers beginning with "X-" have no internet-standard meaning.

- Content-type leads us into the next topic ...

Any headers beginning with "X-" have no internet standard definition. However, certain of them are commonly seen. For example, "X-Mailer:" is pretty much just an advertisement for the software the sender used.

Discussion of Content-type will follow the next sample message.

Sample 2 (part 1)

```
Return-Path: tchen2@rochester.rr.com
Delivery-Date: Sun Apr 28 13:24:50 2002
Received: from heffalump.fnal.gov (heffalump.fnal.gov [131.225.9.20])
        by gungnir.fnal.gov (8.11.6+Sun/8.11.6) with ESMTP id g3SIOoQ13775
        for <crawdad@gungnir.fnal.gov>; Sun, 28 Apr 2002 13:24:50 -0500
  (CDT)
Received: from CONVERSION-DAEMON.smtp.fnal.gov by smtp.fnal.gov
 (PMDF V6.0-24 #37519) id <0GVA00L01J596P@smtp.fnal.gov> for
 crawdad@gungnir.fnal.gov (ORCPT crawdad@fnal.gov); Sun,
 28 Apr 2002 13:24:49 -0500 (CDT)
Received: from mailout5.nyroc.rr.com ([24.92.226.169])
 by smtp.fnal.gov (PMDF V6.0-24 #37519)
 with ESMTP id <0GVA00F63J57LD@smtp.fnal.gov> for crawdad@gungnir.fnal.gov
 (ORCPT crawdad@fnal.gov); Sun, 28 Apr 2002 13:24:45 -0500 (CDT)
```

- So far this is very similar to Sample 1. I don't know anyone in rr.com, but that's a nationwide ISP with wireless service, so it could be someone on the road.
- I see that, as usual, it was originally addressed to crawdad@fnal.gov and routed to crawdad@gungnir.fnal.gov by my forwarding setting on the mail gateway.
- The bracketed IP address really does belong to rr.com and resolves to the given hostname, and vice-versa. This probably is a mail server belonging to the ISP.

The "owner" of an IP address can be checked with

whois -h whois.arin.net 24.92.225.169

If the output refers you to RIPE or APNIC, try again with "ripe" or "apnic" in place of "arin" in the whois server host name.

Sample 2 (part 2)

```
Received: from Fvvm (roc-66-66-65-152.rochester.rr.com [66.66.65.152])
  by mailout5.nyroc.rr.com (8.11.6/Road Runner 1.12) with SMTP id g3SIObH2
  6197 for <crawdad@fnal.gov>; Sun, 28 Apr 2002 14:24:38 -0400 (EDT)
Date: Sun, 28 Apr 2002 14:24:38 -0400 (EDT)
From: tchen2 <tchen2@rochester.rr.com>
Subject: Visibility
To: crawdad@fnal.gov
Message-id: <200204281824.g3SIObH26197@mailout5.nyroc.rr.com>
MIME-version: 1.0
Content-type: multipart/alternative;
 boundary="Boundary_(ID_fXVca0h7nUnO6TLS5mhmhQ)"
```

- From this point on I have no trust in the headers. The originating host called itself "Fvvm" (no domain) and the mail server believed in a different name for that IP address.
- The From address does align with the mail server used to send, but some viral programs are clever enough to do that.
- The Message-id seems to have been provided by the ISP's server, not the originator.

- Another MIME multipart type ...

MIME

- In The Beginning ... all email was plain ASCII text. Lines were short and messages were small.
- People wanted to send
  - Huge messages
  - Non−ASCII characters
  - Binary files, tagged with identifying information
  - Groups of related items.
- MIME was invented to handle all of this, while letting mail pass successfully through unmodifed Transfer−Agents.
- Content−transfer−encoding (quoted−printable or base64) shields non−ASCII content.

Ah, the internet was a much simpler place back then, when there were only forty of us using it ...

Content−Type

- The Content−type header of the message specifies the overall type of the message body and possibly the character set or part separator.
  - text/
    - plain, html, rtf, richtext, enriched, sgml, xml, ...
  - application/
    - postscript, pdf, pgp−encrypted, msword, vnd.ms−powerpoint, x−tar−gzip, octet−stream, ...
  - audio/
    - basic, mpeg, x−wave, ...
  - message/
    - rfc822, external−body, delivery−status, ...
  - multipart/
    - digest, mixed, alternative, related, signed, encrypted, ...

These are not all the top−level content types that exist. See http://www.iana.org/assignments/media−types/index.html for more.

application/octet−stream is a sort of "catch−all" for binary data. Some mailers don’t bother to put in a more specific content−type, leaving it to the file name suffix to convey that information. This is poor practice −− those same mail readers that infer the content type from the suffix allow executables to slip through tagged as audio or image types.

Multipart

- mixed
  - A collection of sub−parts, each with its own type and other tagging information. No interrelationship is assumed.
    - E.g.: text/plain describing enclosure + application/octet−stream
- alternative
  - All sub−parts are presumed to convey the same information, in different formats or with different degrees of fidelity. The mail client should display the best (i.e., last) one that it understands.
    - E.g.: text/plain + application/pdf + application/msword
- related
  - Sub−parts compose a compound object and one of them may be designated the root, referring to others by a Content−id tag.

Klez worm

```
MIME-version: 1.0
Content-type: multipart/alternative;            ⇦ an error?  should be /related?
 boundary="Boundary_(ID_fXVca0h7nUnO6TLS5mhmhQ)"

--Boundary_(ID_fXVca0h7nUnO6TLS5mhmhQ)
Content-type: text/html
Content-transfer-encoding: QUOTED-PRINTABLE

<HTML><HEAD></HEAD><BODY><iframe src=3Dcid:LLH74S35G height=3D0 width=3D0>
</iframe><FONT></FONT></BODY></HTML>  ⇦ next sub-part to be rendered in a 0x0 frame

--Boundary_(ID_fXVca0h7nUnO6TLS5mhmhQ)
Content-id: <LLH74S35G>
Content-type: TEXT/PLAIN; NAME=virus_removed_by_FNAL-Postmaster.txt
Content-transfer-encoding: 7BIT
Content-disposition: attachment; filename=virus_removed      ⇦ foiled!
Content-description: The Original Attachment has been REPLACED

    The original attachment has been removed from this message.
    The attachment was removed because it contained a suspected virus.

--Boundary_(ID_fXVca0h7nUnO6TLS5mhmhQ)

--Boundary_(ID_fXVca0h7nUnO6TLS5mhmhQ)
Content-id: <LLH74S35G>
Content-type: application/octet-stream; name="459820_2_b5gif[2].html" ⇦ bonus track
Content-transfer-encoding: BASE64
```

Maybe what I'm flagging as an error is actually part of the exploit that sneaks the executable content through the preview step of defective mail readers.

Some people have been unable to see that the virus was in fact removed because the iframe directive of the first part concealed the second part.

The third part is completely empty and is interpretted by default as a text/plain.

The fourth part is a file plucked at random from the infected machine's disk. How fun.

What can an individual do?

- For general SPAM problems
  - Delete it.
  - Report it to the originating (or last−reliably−known) ISP.
  - If it's criminal in nature, the US Treasury or other L.E.A.
- Viruses and other nastygrams
  - Receive your mail on an immune platform or through a filtered channel.
  - Keep your A/ V up to date.
- To save yourself a lot of annoyance
  - Install your own mail filter.
    - This can weed out junk *and* pre−sort your important mail. Possibly it can even answer some of your mail for you!
    - For Unix: procmail is the clear winner.

Do not rely solely on the mail gateways as your virus protection. Some things in email may exploit defects in your mail reading software without being considered a virus by the mail gateways.

My advice (not official lab policy): use a web browser for browsing the web; use something else for your email. Use windows to let the sunlight in; use something else for computing.

Procmail Example

- In $HOME/.forward ...
  `"|exec /usr/local/bin/procmail"`
- In $HOME/.procmailrc ...

```
LOGFILE         = $HOME/.procmail.log
LOGABSTRACT     = all
MHDIR           = $HOME/Mail
MAILDROP        = /var/mail/crawdad
# safety net
:0 c
safety-net
:0 ic
| cd safety-net && rm -f dummy `ls -t msg.* | sed -e 1,32d`
# Spim Chee (Korean spam)
:0
* ^Content-type:.*charset=(ks_c|iso-2022-kr|euc-kr)
$MHDIR/spam/.
# Everything else.  First, suppress duplicates ...
:0 hW : msgid.lock
| formail -D 32000 msgid.cache

:0 :
$MAILDROP
```

I started using procmail only very recently. Once you read the manual, even casually, it's not nearly as inscrutable as it seemed.

Yes, you can have the mail gateways forward your mail to your favorite unix system, pass it through procmail there, then forward it back to an IMAP server for later reading. Naturally the people who support email here are not responsible for anything that happens to your email before you get it back onto one of their servers.